Published in final edited form as:
Nat Rev Genet. 2015 March ; 16(3): 146–158. doi:10.1038/nrg3885.

Quantitative and logic modelling of gene and molecular networks

Nicolas Le Novère
Babraham Institute, Babraham research campus, Cambridge CB22 3AT, United Kingdom

Abstract

Behaviours of complex biomolecular systems are often irreducible to the elementary properties of their individual components. Explanatory and predictive mathematical models are therefore useful for fully understanding and precisely engineering cellular functions. The development and analyses of these models require their adaptation to the problems that need to be solved and the type and amount of available genetic or molecular data. Quantitative and logic modelling are among the main methods currently used to model molecular and gene networks. Each approach comes with inherent advantages and weaknesses. Recent developments show that hybrid approaches will become essential for further progress in synthetic biology and in the development of virtual organisms.

Introduction

Even the simplest cellular process involves many molecular components, which display non-linear behaviours and interact in a non-independent and non-additive manner. This complexity hinders any intuitive grasp of the behaviours of a system: for example, a cell, an organ or an organism. Detailed quantitative characterization of each component in isolation is of little help. However, concurrent mathematical descriptions of all the relevant interactions within a system can help to describe its structure, provide understanding of its function and predict its behaviour or 'misbehaviour'. Over the past few decades, MATHEMATICAL MODELS of molecular and gene networks have become an important part of the research toolkit for the biosciences. Their NUMERICAL SIMULATIONS complement experiments aimed at understanding the molecular basis of cell function. They form a unique tool for predicting emergent properties of complex systems. For instance, work on bacterial chemotaxis provides a lasting example of a successful collaboration between modelling and experimental approaches(1,2). Models have increased in size and complexity, culminating in recent efforts such as the comprehensive reconstruction of human metabolism3, the complete model of a microorganism (4) and multiscale models of whole organs (5) and plants (6). Genome-wide reconstructions of gene interactions in various cell types (7,8) mean that we can now model complete gene regulatory networks with increased accuracy.

Mathematical models are constructed and used in a cycle that involves building the structure of the model, choosing mathematical expressions to characterize the networks with

Correspondence to lenov@babraham.ac.uk.

Europe PMC Funders Author Manuscripts

increased accuracy. relationships between its components, finding PARAMETER VALUES and INITIAL CONDITIONS, and performing numerical simulations and other mathematical analyses that can both reproduce observations and lead to predictions (BOX 1). The availability of cheap and easy-to-use computers, coupled with the generation of large amounts of experimental data in digital form, has triggered the development of many methods to model and simulate molecular and gene networks9 (FIG. 1). This diversity of methods has led to uncertainty as to which approach is the most relevant (TABLE 1). QUANTITATIVE MODELS, which are based, in particular, on the application of systems theory to CHEMICAL KINETICS, have been used to describe metabolic networks (10–13), signalling pathways (1,14) and gene regulation (15–18). In addition, the advent of experimental insights of a qualitative nature, such as gene targeting and phenotype screens, has led to the development of methods to model gene regulatory networks on the basis of logic rules (that is, logic, or logical, models) (19,20).

This Review presents some of the approaches available and provides guidance on how to choose the most appropriate one, a choice that depends on the research question and the available data. I focus on dynamic models of biochemical reactions in homogeneous (that is, well-stirred) compartments; this Review does not cover the structural analysis of networks (21) and steady-state analysis (22), nor does it describe methods to spatially model reaction diffusion, which have been covered elsewhere (23,24). Although most of the examples are from gene regulatory networks, all of the methods described can be used to model a wide range of biological processes.

Four network representations in systems biology

Network representation and analysis sit at the core of systems biology, and behind each mathematical model of molecular or genetic processes is a network. Many network representations have been used, and although they might seem to form an unstructured continuum, one can classify them into four families (FIG. 2) . These representations are more than mere illustrations; they convey deep semantics about the underlying biological process together with the context of the study and the hypotheses made. It is important to choose the appropriate representation on the basis of the question asked and the data available. The choice of network representation must be made early in the modelling process because it affects the selection of the modelling and simulation methods, as well as the processing of the data used to build and validate the models.

Interaction Networks

Interaction networks (FIG. 2a) are used to represent lists of physical or functional (for example, genetic) interactions. Interactions are often undirected (that is, if X interacts with Y, Y interacts with X), and the graph is non-sequential (that is, one cannot start on a node and draw a path through the map via successive arcs). A subset of interaction networks have directed but ‘unsigned’ arcs: we know that X affects Y, but not whether it activates or represses it. Such networks can be classified as either interaction networks or activity flows (see below). Gene and protein interaction networks have been reconstructed to obtain a comprehensive view of genome regulation 8,25–27 or to understand specific processes (for instance, the regulation of pluripotency) 28–31. Although interaction networks are useful for analysing the structure of systems or the results of their perturbation21, their lack of

mechanistic insights and their static nature make them unsuitable for representing dynamic models.

Activity flows

Often, the precise molecular mechanisms underlying the effect of a mutation or a chemical perturbation are unknown, with the only information being that X is increased or decreased. Activity flows (also known as influence diagrams) allow this information to be represented in a concise way (FIG. 2b). In other words, activity flows are used when the detail of a chemical reaction is not known or is not considered key to understanding the biology. This is often the case in the representation of signalling pathways or gene regulatory networks. For instance, the non-metabolic parts of the Kyoto Encyclopedia of Genes and Genomes pathway (KEGG PATHWAY) database and Science Database of Cell Signaling use activity flows. Some examples of gene networks include maps of sea-urchin development from the Davidson laboratory (see Davidson's maps) and transcription networks in stem cells (32), neural development (33) and neurons (34). The main nodes are activities, and they are linked by arcs representing the direction of the influence. Activity flows are therefore suitable for representing the transfer of information. They have been standardized with the Systems Biology Graphical Notation (SBGN) activity flow language. Although these maps are directed and sequential, one cannot infer a mechanism behind an edge. The statement "X activity stimulates Y activity" can refer to a wide variety of mechanisms, including activation of Y production, inhibition of Y degradation or stabilization of a high-activity state of Y. Because of the qualitative nature of the information provided, activity flows are the natural representations for qualitative models and, in particular, for logic models.

Process descriptions

Process descriptions (FIG. 2c) are bipartite graphs with two types of nodes: the variables whose evolution one wants to follow; and the processes that decrease or increase (consume or produce) the values of these variables. The arcs of process description maps are directed, and the networks are sequential. Process descriptions are suitable for representing transfer of mass. They have long been used to describe biological systems and represent an evolution of the chemical reaction network that was present in the first metabolic maps (35,36). Process descriptions used in biochemistry have been standardized with the SBGN process description language. The granularity of description allows mechanistic descriptions, making process description maps suitable representations of chemical kinetic models. Unfortunately, this granularity comes at a cost. In contrast to the statements in entity relationship maps (see below), the processes are not independent and lead to a combinatorial explosion. For instance, a promoter that binds to a transcription factor will exist in four states: bound, unbound, methylated and unmethylated. Using it in another reaction — for example, in binding to a polymerase — requires four processes. This combinatorial explosion also affects the corresponding chemical kinetic models. Process descriptions are used widely to depict metabolic processes, whether central metabolism or metabolic reactions associated with signalling or gene regulation. Accordingly, the metabolic networks in the KEGG pathway database (37) are described in process descriptions, as are pathways in the Reactome pathway database (38).

Entity relationships

In entity relationships (FIG. 2d), one represents entities (for example, a gene), statements about those entities (for example, an interaction or a methylation status) and the influence of entities on statements (for example, the stimulation of an interaction). Entity relationship maps introduce the directionality of influences (that is, “X stimulates Y” is different from “Y stimulates X”) and offer a granularity of representation that is suitable for molecular mechanisms (39). Entity relationships have been standardized using the SBGN entity relationship language (40), and such maps have been constructed to represent molecular events underlying, for instance, the cell cycle (41,42) and apoptosis (43). The maps are built through the accumulation of independent relationships, each of which describes a fact (for example, a site is phosphorylated or phosphorylation is stimulated). Entity relationships are thus a perfect graphical representation for rule-based models (44,45). Although interest in rule-based models is growing46 and they represent an interesting path for further investigation, they are still not in mainstream use. Furthermore, as their current use is mostly centred on signalling pathways, I do not describe them further in this Review.

Mining information to build models

As described above, network visualizations such as process descriptions and activity flows represent the pathway counterparts of individual modelling approaches, such as chemical kinetics and logic modelling, respectively. It is important to understand that each type of representation, and hence its corresponding modelling approach, is best suited to different situations and will provide different insights (TABLE 1). A key factor in the choice of representation is the type of knowledge available about the system: do we know only the direction of regulations, or are the mechanisms underlying the regulation elucidated? Moreover, the nature of available experimental data is important: are quantitative timecourse experiments available that yield data on concentrations or gene expression levels? Or can we use only phenotypes and normalized measurements?

Depending on the system being modelled, a bottom-up or knowledge-based approach can be adopted, hereby information on the components to include and their relationships is obtained from scientific literature or public databases that contain previously generated models or information to incorporate into building blocks. By contrast, it is sometimes possible to infer starting points for building models directly from experimental data sets. This is called a top-down or data-based approach.

The knowledge-based approach

The existing corpus of models is often the most useful source of information when starting to build a mathematical model. Even if no models have been created specifically for a particular problem, one may be able to reuse models, or parts of models, created to answer others. For instance, a model of epidermal growth factor receptor (EGFR) signalling in tumour progression (47) was built using models of phosphoinositide 3-kinase (PI3K)48, mitogen-activated protein kinase (MAPK) (49) and Janus kinase (JAK)–signal transducer and activator of transcription (STAT) signalling (50). Although descriptions of mathematical models in the literature vary widely, it has become commonplace to provide the model

source as supplementary information, and the development of standards such as SYSTEMS BIOLOGY MARKUP LANGUAGE (SBML) (51) permits the reuse of these descriptions in different software. The models are also often deposited in public databases. This makes it possible to search and retrieve relevant models using various criteria (for example, biological process, biochemical component, organism and authors). BioModels (32) is an example of a database that provides a large collection of mathematical models of biological processes encoded in SBML.

Nevertheless, existing models developed to answer a specific question in a given context are rarely directly reusable. Furthermore, in some areas of cellular biology, computational models are scarce. Therefore, another approach is to retrieve building blocks from biological databases, a comprehensive list of which can be found in the annual database supplement of the journal Nucleic Acids Research (52). Of particular use are databases that list metabolic and signalling pathways, or enzymes. After listing the molecular components to include in the model, one must determine physical or genetic interactions between them (BOX 1). There are many protein interaction databases, distributing information generated with methods such as immunoprecipitation or yeast two-hybrid assays. One of the largest resources is the well-curated IntAct (53). In addition to physical molecular interactions, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (54) distributes functional interactions inferred from various sources, such as coexpression and text mining. However, the sole interactions cannot be directly used in dynamic models.

Pathway databases provide functional information such as reactions or regulations. The most frequently used is KEGG PATHWAY (37) , which provides a large collection of manually drawn maps linked to underlying databases. Reactome (38) provides the deepest level of curation, albeit with a more limited coverage. Its maps exported in SBML can be used directly as a starting point for further modelling. The BioCyc database collection, which includes the metabolic pathway database MetaCyc (55), provides the largest coverage in terms of both organisms and biological processes. Metaresources such as ConsensusPathDB (56) and Pathguide (57) provide access to a large variety of pathway databases. Quantitative information about biochemical reactions can also be found in databases such as SABIO-RK (58) and BRENDA (59). The Path2Models project (60) is an example of large-scale generation of mathematical models from pathway and biochemistry databases. With more than 140,000 annotated SBML files covering all domains of cell biology, it provides starting points for mathematical models.

The data-based approach

An alternative to reconstructing molecular and gene network models from known interactions is to infer their topology from experimental datasets. Most inference software was originally developed to infer the network from gene expression data (61,62), although information can be extracted from various molecular phenotypes (63). Such BIOLOGICAL NETWORK INFERENCE is part of reverse engineering (64) (another part is parameter estimation, which is discussed below). Comprehensive surveys of existing network inference methods and tools have been undertaken elsewhere (65,66).

Different approaches perform better with different kinds of measurements (for example, steady state or time courses) and provide different types of information (for example, directed versus undirected arcs) (FIG. 3). Statistical methods such as correlation (for instance, WGCNA) and regression (for example, TIGRESS (67)) reveal whether variables are independent. For instance, if two genes are consistently upregulated or downregulated together, the chances are that they share some regulatory features. A recent example of such an analysis is the definition of a gene regulatory network that controls naive pluripotency (68). Information theoretic methods such as mutual information — for instance, context likelihood of relatedness (CLR) (69) and ARACNE (70) — express how much information one gets on a variable value when the value of another variable is known; for example, how much more certain we are of the expression of a gene when we know the expression of another one. Correlation and information theoretic methods infer networks with undirected arcs, whereas regression methods can predict directed influences. Undirected networks cannot be easily used to build dynamic models, and some methods have been proposed to add directionality to interaction networks (71). Directed influences are also predicted by probabilistic methods such as Bayesian inference (72) — for example, Banjo (73) and CatNet — which compute the probability that a certain set of data is produced by various networks. One can then select the most probable network. Such methods can infer causal networks from large-dimensional data sets, such as signalling pathways from multi-parametric flow cytometry (74). Bayesian inference methods permit the construction of directed, sometimes signed (in which arcs represent positive or negative effects), networks from functional genomics data sets and interactomes. Thereafter, those networks can be transformed into fully fledged logic models (for instance, with CellNetOptimizer (71)). Bayesian inference networks are graph models. Other methods based on graph theory have recently been proposed (75). Finally, methods based on ordinary differential equations (ODEs) (76) (for example, Inferelator (77) and NIR (78)) are particularly suited to time-course data and aim at inferring quantitative and dynamic interactions between genes and molecules. Only in the case of linear models can interactions and their signals be easily inferred from data (79,80). Methods can be compared and ranked between and within any of the categories above. However, the advantages and pitfalls of the various categories are often complementary, and the best approach to robustly infer accurate networks seems to be a combination of methods (81). Recent software tools can be used to help to implement such a multiple-prong strategy (82).

Quantitative kinetic models

The most common approach used to model molecular networks is based on the application of systems theory to chemical kinetics. A state variable of the model represents the quantity of a molecular species (such as a metabolite concentration), an amount of mRNA or the activity of a gene. This quantity is dynamically controlled by the combination of all processes that increase the level of the molecular species (for example, synthesis, import and activation) and all processes that decrease the level of molecular species (for example, degradation, export and inhibition). Each process is characterized by a rate that can be modulated by various parameters, including the quantity or activity of other molecular species. A more thorough explanation of the basis of chemical kinetics can be found elsewhere (83,84). The system can be simulated by computing the changes in variable

values over small intervals of time. This is done either by reconstructing ODEs that represent the whole system or by simulating each process separately using, for instance, stochastic simulation approaches (BOX 2). A plethora of tools are available to create such models and simulate their behaviours (TABLE 2). These programs range from simple simulators of models encoded in SBML, such as SBMLsimulator (85), to fully featured modelling environments, such as COPASI (86). A good starting point for exploring the current offerings is the SBML Software Guide.

Depending on the size of the model, the type of information available and the granularity of answers sought from the simulations, different approaches can be used to represent the regulatory mechanisms (87). Used in conjunction with quantitative experimental data, such models are powerful tools for decrypting and understanding systems. Testing the effect of mutations in silico and replacing them within the context of complete pathway models can direct experimental perturbations and help to interpret their result, as shown by the discovery of a new PI3K-insensitive activation of mammalian target of rapamycin (mTOR) by insulin (88). Their simulations provide quantitative and temporal predictions (FIG. 1) , which can be crucial for understanding biological processes. For instance, quantitative models of nuclear factor-κB (NF-κB) signalling predicted oscillations that were shown to be essential to NF-κB-dependent transcription (89). It was later shown with experiments and stochastic simulations of single cells that a complex model of three feedback loops accounted for the response to pulsatile stimulations with different frequencies of NF-κB signalling and patterns of NF-κB-dependent transcription (90).

Process description models can be developed with varying degrees of mechanistic insight. I present only three of them below, going from more biochemically accurate representations to more abstract descriptions.

Chemical kinetics

Chemical kinetics. Biochemical and cellular phenomena follow the laws of chemistry and physics. If the underlying chemical processes are known, the behaviour of the system can be described on the basis of thermodynamics and chemical kinetics. The rate of elementary reactions is determined by the relative activities (molecular concentrations or gene activity) of the reactants and products, and by rate constants, which are themselves linked to the free energy of the different states. The most generic representation is based on the MASS ACTION LAW, which takes into consideration all of the elementary binding events, dissociation events, catalyses and state transitions driven by the laws of chemical kinetics (91). This approach is the most accurate if one wants to quantitatively explore the behaviours of simple molecular (92) or gene regulatory networks (90). BOX 2 presents an example of such a model. However, if some conditions are met, simplified rate law can be used (93). The most frequent simplification is HENRY–MICHAELIS–MENTEN kinetics, which allows the enzyme–substrate complex (or ligand–receptor or promoter–transcription factor complexes) to be ignored if the concentration of this complex does not change or if the rate of binding is much faster than the rate of product formation. These rules can be generalized to represent a large range of regulatory processes, including non-independent regulations (94,95).

However, the detailed molecular mechanisms underlying a process are often unknown, or the aim is to reduce their complexity and concentrate on the main effect of a component. In such cases, the model needs to be derived from experimental observations, for which several approaches have been developed. Below, I describe in detail the use of Hill functions; an example of an alternative approach is the use of S-systems (11,96).

Hill functions

Hill functions. A generic way to represent modulations if the underlying biochemical mechanisms is unknown is to multiply the default activity of a gene or the velocity of a process by Hill functions, of the form $X^{\alpha} / (X^{\alpha} + K^{\alpha})$. K represents the activity of X for which the effect is 50%. If α is positive, the value of the function is 0 for X = 0 and 1 when X is very large. Therefore, X is an activator. If α is negative, the value of the function is 1 for X = 0 and 0 when X is very large. Therefore, X is a repressor. K shifts the response curve horizontally and represents the sensitivity of the response to X. The exponent α controls the steepness of the response to X (that is, the cooperativity). With $\alpha = 1$, the systems respond linearly to X when X is not very large in comparison to K. As α increases, the response becomes ever closer to a threshold with a small dynamic range. It is easy to see that one can then combine the effects of different regulators by multiplying the Hill function terms. One can even combine different Hill functions for the same regulator to represent several modulations at different concentrations (97). More-complex representations can take into account basal expressions in the absence of activators, the non-independence of regulators, and so on.

This approach has been useful in modelling many biochemical systems, such as calcium signalling (97), the cell cycle (98) and oscillators in general (99). In the field of gene regulatory networks, Hill functions have been useful for understanding the control of segmentation (16,100,101). Although Hill functions and similar phenomenological descriptions are extremely useful and practical, one must keep in mind that they rely on assumptions that might not always be valid. For instance, a core hypothesis is that the binding and dissociation of the regulator are extremely fast in comparison to the process regulated and can therefore be ignored because the fraction of bound regulator is at equilibrium. Although this hypothesis does not always hold for signalling cascades — in which perturbations and responses are on the same time scales as association and dissociation, thus resulting in distorted kinetics — it is generally considered to be valid for gene regulation because of the considerable difference between the dynamics of transcription factor diffusion and binding in comparison to transcription and translation.

Piecewise linear differential equations

A further approximation is taking the limit of Hill functions by using step functions (also known as Heaviside functions). In this framework, the response to modulators (that is, their effect on the rates of the processes) is discrete: for instance, 0 below a certain amount of modulators and maximal over it. These approaches, which were introduced in the 1970s (102), have been intensely studied and improved over time. Further simplifications were recently introduced, such as the use of discrete instead of continuous time, in which the system is updated at regular intervals (103) . Nevertheless, their use in quantitative kinetic

models has remained limited, although a subset was used to model gene regulatory networks that also consider qualitative descriptions of variables (104). Software such as the Genetic Network Analyzer (105) allows the construction of such models and the study of their possible stable states. The generation and analysis of such qualitative models is similar to the study of logic models (see below).

Parameterization of quantitative models

A bottleneck in building quantitative kinetic models is the lack of suitable parameters, such as rate or equilibrium constants. One way of addressing this issue is to estimate the values of those parameters using experimental data sets. The estimation of model parameter values is a form of GLOBAL OPTIMIZATION and part of the reverse engineering of molecular and gene networks (64), a complement of network inference presented above. The principle is to test different parameter values and select the sets that minimize an error function. This function can be derived, for instance, from the difference between the values of model variables and their experimental determination. As testing all of the possible data sets is impossible, the difficulty of the procedure is sampling the parameter space and selecting the next set to test on the basis of past values of the error function. Many methods have been developed106, and several of them have been implemented in software such as COPASI86. It is important to note that the experimental data do not necessarily directly correspond to the variable of the model. As far as there is a mathematical transformation that can lead from the experiment to unique values of the model variables, parameters can be estimated. A model for which one can theoretically find values for unknown parameters from adequate data is known as an IDENTIFIABLE MODEL (64).

Limitations of quantitative kinetics modelling

Despite quantitative kinetic modelling being a natural representation of molecular and gene networks, and despite the approach providing the most precise predictions, the lack of kinetic data (and of quantitative data in general) hampers its use in many situations. Although central metabolism has been characterized quantitatively for more than a century, little is known about reaction kinetics or equilibrium constants in the realm of gene expression or signalling. Moreover, reaction-based descriptions are sensitive to the existence of multiple-state entities — for instance, proteins with several conformations, covalent modifications, methylation states and promoters with different occupancies — or multiple-component complexes. In such situations, process descriptions lead to a combinatorial explosion of both variables and processes, as described above. Finally, the theoretical framework of chemical kinetics assumes a homogeneous distribution of participants in volumes. Most signalling reactions involve a few molecular partners that are heterogeneously located in spatially complex cellular compartments, such as membranes. Gene regulation involves even fewer partners, and reactions take place in a complex spatial domain composed of folded nucleic acids. These limitations are tentatively addressed by other approaches, such as logic modelling.

Europe PMC Funders Author Manuscripts

Logic models

Logic models are characterized by the assignment of new values to model variables on the basis of the result of logical statements (107). These statements combine the values of the model variables. For instance, C = x if A = y and B = z. Logic models are most often used in conjunction with qualitative variables that are represented by a few discrete integers. In most variants, time is not represented in the model or during its simulation. State transitions take place at each time step; however, a time step can represent a different duration for different transitions. Once a logic model is built, one can produce trajectories (pseudo time courses) and study the possible ATTRACTORS of the system (BOX 3).

Logic models are versatile: a variable can represent almost anything, such as a gene activity, the presence of a protein or the state of a cell. They are flexible: the state of a given cellular component can be represented by one or more variables, with different sets of values. For instance, EGFR could be represented by a single variable: EGF would switch EGFR 'on', and once on, EGFR would switch its targets on. Alternatively, if we want to model separately the effect of drugs that target the binding of EGF (such as cetuximab) and those that block the tyrosine kinase activity of EGFR (such as gefitinib), we can represent EGFR as two variables: EGFRbinding would be switched on by EGF and 'off ' by cetuximab. Once on, it would turn on the variable EGFRtk, which is itself turned off by gefitinib. Finally, perturbations, such as the effects of inhibitors and mutations, can be tested straightforwardly in a logic model.

Logic modelling remained fairly theoretical until the end of the 1990s, when it aided the modelling of gene regulatory networks involved in the regulation of development (108,109). Since then, logic models have played an important part in increasing our understanding of cell differentiation; recently, such models have been used in the study of haematopoiesis (110) and embryonic stem cells (29). In a similar way to the modelling of gene regulation, the modelling of signalling pathways suffers from the lack of kinetic information. Logic modelling has thus made a difference in, for instance, our understanding of the pathways underlying cell fate in cancer (111–113).

Different types of logic models

Different types of logic models. Logic modelling of biological systems is a rapidly expanding field with the development of new methods (107). In particular, the representation of time is the subject of many variations.

An updating scheme is needed when simulating logic models; variables in a logic model can be updated synchronously, with the values of all variables being calculated after each transition, or asynchronously, when variables undergo these transitions one at a time (114). This requires mechanisms of prioritization and delay (115). Probabilistic Boolean networks allow greater flexibility by providing alternative logical functions with different probabilities for updating a node (116). Stochastic simulations of such models allow biological noise and the resulting variability of responses to perturbations to be taken into account (117,118). In the most frequent variant of logic models, Boolean models, variables take the values 0 or 1 (119). The use of multi-valued variables allows the encoding of much

richer behaviours with alternative influences of one variable on another on the basis of its state. Examples are semi-quantitative proportionality, with off, low and high activities resulting in no, weak and strong effects, respectively, and alternative signs for the resulting influences with off, low and high activities resulting in no effect, stimulation and inhibition, respectively. Continuous regulations were ultimately introduced in logic models with FUZZY LOGIC (120). The limit of this trend is reached through methods that allow the transformation of logic models into quantitative ones using, for instance, Hill functions (111).

As is the case for quantitative kinetic models, a crucial part of building logic models is the parameterization, which might consist of deciding on the terms (also called ‘gates’ or ‘connectors’) of the logical function regulating the values of the variables. Advanced software (121) can provide different approaches for fitting logic models with experimental information. The adequacy of data fitting to the variants of logic modelling is discussed elsewhere (122). Logic models are also easily amenable to model checking (123), a set of approaches that seek to evaluate whether a model produces a given behaviour.

Limitations of logic models

Despite their ease of use, logic models present a few inconveniences. First, although scaling up logic models is relatively easy, the number of states increases exponentially with the number of variables. Computing state-transition diagrams and attractors (that is, stable solutions of the model) is not necessarily a problem with modern computers. However, gathering insights from these analyses becomes difficult. Second, the lack of a representation of time in simple logic modelling approaches makes it difficult to take into account slow and fast processes and delays. Third, because of its purely qualitative nature, it can often be difficult to choose between alternative behaviours proposed by logic modelling: for instance, a functional negative feedback would always lead to a periodic behaviour, with the model cycling between states, whereas in a quantitative kinetic mode it would lead to either equilibrium or oscillation, depending on the strength of the feedback. To alleviate such shortcomings, more-complex logic model analysis have been designed, some of which are implemented in the software listed in TABLE 2.

Towards modular hybrid models

So far, models have been developed mostly in a monolithic manner; that is, a system is described using a single model, based on a single modelling method. This design approach is reaching its limit, and a paradigm shift is needed to support the emerging fields of research. Synthetic biology involves the assembly of existing parts (‘biobricks’) to create new systems. Predicting the behaviour of these systems requires a model that is built by merging models of individual parts. Systems pharmacology bridges the fields of pharmacometrics and systems biology (124) , and requires statistical pharmacokinetic and pharmacodynamic models that can ‘talk’ to mechanistic network models. Finally, the development of virtual organs and organisms (single-cell or multicellular) relies on the assembly of many different models of processes taking place at different scales (5,6).

The next generation of models needs to be modular, in which different processes are modelled independently and the integration of models takes place during simulation through

variable transformation and synchronization. Such modularization of models and simulations would allow more even distribution of the model-development burden, more efficient version handling and the use of different simulators for the relevant modules. Chemical kinetics is intrinsically modular (that is, each reaction can be seen as a module), and combining different quantitative simulation algorithms was proposed by members of the E-Cell Project (125), in which separate processes can be simulated using different 'steppers' that update the system when synchronization is needed.

Furthermore, because of the range of biological processes to represent and the heterogeneity of experimental measurements available to build and validate the models, it is necessary to use hybrid modelling approaches, in which qualitative and quantitative representations are used in the same model. This approach was proposed in 1995 for the modelling of gene regulatory networks(126). One possibility is to use ODEs and piecewise linear differential equations to represent the evolution of different quantities (127). Logic models (for instance, of signalling pathways) can also be combined with quantitative models (for instance, of metabolism) by generating kinetic representations of the logic parts when necessary (128) (that is, 'ODEfication' (111)). The awareness that biological processes take place on different time scales led to the development of models that use iterative quantitative steady-state representations of metabolism, in which the iteration is coupled by logic modelling of gene regulatory networks (129–131).

This approach culminated in a complete whole-cell model of *Mycoplasma genitalium* that combines modularization and hybrid modelling. This model used a mix of ODEs, stochastic processes and flux balance analysis to simulate 28 modules representing gene networks, signalling and metabolism (4) . Recent work also showed that synchronizing several simulators allows the concurrent use of different representations and simulation procedures. Such synchronization can be applied, for example, to a whole-neuron model, in which chemical kinetics is used to model synaptic signalling and cable approximation is used to model electrical signals at the level of the entire neuron (132).

Conclusions

Using mathematics to model and understand the world is one of the cornerstones of science. With the rise of systems biology at the end of the twentieth century, the adoption of mathematical modelling has been rapid in genetics and molecular biology. The launch of a large number of projects by many institutions led to a demand for scientists with good knowledge of molecular and cellular biology and an understanding of the modelling process, along with its underlying mathematics. However, there is a skill shortage in the mathematical modelling sector; furthermore, the information and training needed to address this skills gap is spread across disciplines and institutions. Typically, the community of quantitative kinetic modelling originates from physics and engineering, whereas scientists who develop and use formal modelling have a background in bioinformatics and/or mathematics.

Mathematical modelling of molecular and gene networks is an important part of systems biology, and numerous methods and models are continually being developed by a vibrant

community. Well-designed, experimentally validated models help us to understand molecular and cellular processes and can predict the effects of drugs or mutations. A greater awareness of the different modelling methods and how to combine them will make such models more versatile and more useful, and new training programmes must strive to encompass all aspects of the modelling process.

Acknowledgements

The author is grateful to Laurence Calzone, Martina Froehlich and Linda Koch, for extensive reading and correction of the manuscript, to Len Stephens and Stuart Edelstein for their corrections and advice, and to Claudine Chaouiya, Pedro Mendes, Julio Saez-Rodriguez and Dennis Thieffry for help with the bibliography.

Further information

ARACNE: http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/ARACNE

Banjoo: https://www.cs.duke.edu/~amink/software/banjo/

BIOCHAM: http://lifeware.inria.fr/biocham/

BioCyc: http://biocyc.org/

BioModels: http://www.ebi.ac.uk/biomodels

BoolNET: http://cran.r-project.org/web/packages/BoolNet/

BRENDA: http://www.brenda-enzymes.org/

CellDesigner: http://www.celldesigner.org

CatNet: http://cran.r-project.org/web/packages/catnet/

CellNetOptimizer: http://www.cellnopt.org/

COPASI: http://copasi.org

ConsensusPathDB: http://consensuspathdb.org/

Davidson's maps: http://sugp.caltech.edu/endomes

DBSolve: http://insysbio.ru/en/software/db-solve-optimum

E-Cell Project: http://www.e-cell.org/

GINsim: http://www.ginsim.org/

Genetic Network Analyzer: http://www-helix.inrialpes.fr/gna/

iBioSim: http://www.async.ece.utah.edu/iBioSim/

IntAct: http://www.ebi.ac.uk/intact

Europe PMC Funders Author Manuscripts

Inferelator: http://bonneaulab.bio.nyu.edu/networks.html

KEGG PATHWAYS: http://www.genome.jp/kegg/pathway.html

Kohn's maps: http://discover.nci.nih.gov/mim

MetaCyc: http://metacyc.org/

NAIL: http://sourceforge.net/projects/nailsystemsbiology/

NIR: http://dibernardo.tigem.it/softwares/network-inference-by-reverse-engineering-nir

Path2Models: http://www.ebi.ac.uk/biomodels-main/path2models

Pathguide: http://www.pathguide.org/

Reactome: http://www.reactome.org

SABIO-RK: http://sabio.villa-bosch.de/

SBMLsimulator: http://www.ra.cs.uni-tuebingen.de/software/SBMLsimulator/

SBML Software Guide: http://sbml.org/SBML_Software_Guide

Science Database of Cell Signaling: http://stke.sciencemag.org/cm

STRING: http://string-db.org

TIGRESS: http://cbio.ensmp.fr/tigress

WGCNA: http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/

XPP-Aut: http://www.math.pitt.edu/~bard/xpp/xpp.html

Glossary

ATTRACTORS	Stable behaviour of a system, as reflected by a fixed trajectory in the space of all possible states of the system. Examples of attractors are periodic behaviours (for example, oscillations) and steady-states.
BAYESIAN INFERENCE	A method of inference using Bayes' theorem to evaluate the probability of a network given a data set, as a function of the probability that this network produces the data set, the chance probability of this network and the chance probability of the data set.
BIOLOGICAL NETWORK INFERENCE	A procedure whereby an unknown set of biological interactions and processes is deduced from the molecular phenotypes it

	produces: for instance, a list of gene expression, of molecular concentrations or of phenotypes on perturbation.
BIPARTITE GRAPHS	Graphs that contains two types of nodes, in which nodes of one type are only connected to nodes of the other type. For example, in a metabolic network, nodes representing biochemical species connect to nodes representing reactions.
CHEMICAL KINETICS	The study of rates of chemical processes and how they affect the evolution of chemical compounds in a system.
GLOBAL OPTIMIZATION	A branch of numerical analysis that deals with the global optimization of a function or a set of functions according to some criteria. Examples of global optimisation problems in biological network modelling are parameter estimation and flux balance analysis.
HENRY-MICHAELIS-MENTEN KINETICS	A kinetic scheme used in enzymatic reactions. If the formation of an enzyme–substrate complex is faster than the formation of the enzymatic product or if the concentration of enzyme–substrate complex is constant, one can explicitly avoid representing the enzyme– substrate complex. The rate of formation of the enzymatic product is then proportional to the fraction of enzyme bound to the substrate: that is, [E] × [S] / (Km + [S]), where Km is the concentration of substrate necessary to achieve half the maximal reaction velocity.
IDENTIFIABLE MODEL	A model in which the values of its parameters can be unambiguously determined by the data sets available. A model is non-identifiable if alternative sets of parameter values can fit the datasets.
INFORMATION THEORETIC METHOD	Inference methods based on the information theory. Variables (nodes) are linked in a network if information about one variable (for instance, the distribution of its values) is affected by the knowledge of the values of the other.
INITIAL CONDITIONS	Values for the model variables at the start of numerical simulations. These initial conditions might affect the simulation results — for instance, in the case of systems with several stable states that can be reached from different trajectories.
LOGIC MODELS	Mathematical models in which the discrete values of variables are determined by logical combinations of the values of other variables.
MASS ACTION LAW	A law stating that the velocity of a reaction is proportional to the concentration of the reactants it consumes raised to the power of their stoichiometry. For instance, the rate of a reaction

	consuming two molecules of A and one molecule of B will be proportional to $[A]^2 \times [B]$.
MATHEMATICAL MODELS	Descriptions of a system using mathematical concepts and language. Models are composed of a set of variables and a set of equations that establish relationships between the variables.
NUMERICAL SIMULATIONS	Reproductions of the behaviour of a system, obtained by iteratively computing the values of variables in a mathematical model over a certain number of time steps.
OPEN STANDARDS	Standards that are publicly available and that can be implemented without restriction by licensing terms. In computational biology, open standards are additionally developed by the community, and implementations are not subjected to fees.
ORDINARY DIFFERENTIAL EQUATION	[ODEs] Equations describing the change of a variable in a system over time as a function of the values of other variables and parameters in the system. In a model of a biochemical system, the ODEs are derived from the combination of the different processes in which the entity represented by the variable is involved.
PARAMETER VALUES	The temporal evolution of model variables (for example, protein concentrations) is affected by the values of other variables and by parameters such as dissociation constants, kinetic rate constants and reaction orders. Parameter values affect the dynamic behaviour of model variables.
QUANTITATIVE MODELS	Mathematical models in which the values of the variables are determined by numerical analysis of the variable and parameters in the system.
REACTION ORDER	The order of a reaction for a given reactant is defined as the exponent to which its concentration is raised in the rate law that characterizes the reaction. In the case of reactions taking place in a well-stirred, diluted medium, the reaction order of a molecular species is equal to its stoichiometry for this reaction.
S-SYSTEMS	Modelling approaches for biochemical systems in which the creation and destruction of molecular species are expressed as products of the concentration of all the molecular species in the systems raised to a phenomenological order (obtained by fitting the model to experimental data).
STOCHASTIC SIMULATION	Simulation of a model in which each process has a certain probability to occur. Examples of stochastic simulations are solutions of stochastic differential equations, in which noise

	factors are added to otherwise deterministic ordinary differential equations, and dynamic Monte Carlo simulations in which reaction rates are sampled from distributions.
SYSTEMS BIOLOGY GRAPHICAL NOTATION	[SBGN] A set of standardized symbols to represent the entities included in a biochemical network and their relationships. The notation is formed of three complementary languages to represent activity flows, entity relationships and process descriptions.
SYSTEMS BIOLOGY MARKUP LANGUAGE	[SBML] A format to encode mathematical models that is used in systems biology. Although initially focused on non-spatial, reaction-based biochemical models, the language now features packages covering different modelling approaches. SBML is supported by software libraries in different programming languages and can be imported or exported by hundreds of modelling and simulation tools.
FUZZY LOGIC	Approximate logic computation in which the variables can have partial truth values ranging from 0 (false) to 1 (true).

References

1. Bray D, Bourret R, Simon M. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. Mol. Biol. Cell. 1993; 4:469–482. [PubMed: 8334303]
2. Tindall MJ, Gaffney E. a, Maini PK, Armitage JP. Theoretical insights into bacterial chemotaxis. Wiley Interdiscip. Rev. Syst. Biol. Med. 2012; 4:247–59. [PubMed: 22411503]
3. Thiele I, et al. A community-driven global reconstruction of human metabolism. Nat. Biotechnol. 2013; 31:419–425. [PubMed: 23455439]
4. Karr JR, et al. A whole-cell computational model predicts phenotype from genotype. Cell. 2012; 150:389–401. [PubMed: 22817898] Modular model of an entire *Mycoplasma genitalium* cell, including the expression of all genes, all metabolites and signalling pathways. The model is simulated with an hybrid approach, using stochastic simulations, ODEs and flux balance analysis.
5. Schliess F, et al. Integrated metabolic spatial-temporal model for the prediction of ammonia detoxification during liver damage and regeneration. Hepatology. 2014:2–44. doi:10.1002/hep.27136. [PubMed: 24677161]
6. Chew YH, et al. Multiscale digital Arabidopsis predicts individual organ and whole-organism growth. Proc. Natl. Acad. Sci. 2014:1–10. doi:10.1073/pnas.1410238111. [PubMed: 25197087]
7. Gerstein MB, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012; 489:91–100. [PubMed: 22955619]
8. Neph S, et al. Circuitry and dynamics of human transcription factor regulatory networks. Cell. 2012; 150:1274–86. [PubMed: 22959076]
9. De Jong H. Modeling and simulation of genetic regulatory systems: a literature review. J. Comput. Biol. 2002; 9:67–103. [PubMed: 11911796]
10. Chance B, Greenstein DS, Higgins J, Yang CC. The mechanism of catalase action. II. Electric analog computer studies. Arch. Biochem. Biophys. 1952; 37:322–339. [PubMed: 14953444]
11. Savageau MA. Biochemical Systems Analysis III. Dynamic solutions using a power-law approximation. J Theor Biol. 1970; 26:215–226. [PubMed: 5434343]
12. Kacser H, Burns J. The control of flux. Symp. Soc. Exp. Biol. 1973; 27:65–104. [PubMed: 4148886]

13. Joshi A, Palsson BO. Metabolic dynamics in the human red cell: Part I—A comprehensive kinetic model. J. Theor. Biol. 1989; 141:515–528. [PubMed: 2630803]
14. Goldbeter A, Koshland D. An amplified sensitivity arising from covalent modification in biological systems. Proc Natl Acad Sci U S A. 1981; 78:6840–6844. [PubMed: 6947258]
15. Arkin A, Ross J, Mcadams HH. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in phage lambda-infected Escherishia coli cells. Genetics. 1998; 149:1633–1648. [PubMed: 9691025]
16. Von Dassow G, Meir E, Munro E, Odell G. The segment polarity network is a robust developmental module. Nature. 2000; 406:188–192. [PubMed: 10910359] Paper presenting a dynamical quantitative model of the segment polarity gene network in *Drosophila*, and systematically studying effects of parameter value changes. It concludes that the model is robust versus any of the parameters, but only a tiny fraction of the entire parameter space leads to expected results.
17. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000; 403:335–8. [PubMed: 10659856]
18. Gardner T, Cantor C, Collins J. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000; 403:339–342. [PubMed: 10659857]
19. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 1969; 22:437–67. [PubMed: 5803332]
20. Thomas R. Boolean formalization of genetic control circuits. J. Theor. Biol. 1973; 42:563–85. [PubMed: 4588055]
21. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 2004; 5:101–13. [PubMed: 14735121]
22. Bordbar A, Monk JM, King Z. a, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nat. Rev. Genet. 2014; 15:107–20. [PubMed: 24430943]
23. Takahashi K, Arjunan SNV, Tomita M. Space in systems biology of signaling pathways--towards intracellular molecular crowding in silico. FEBS Lett. 2005; 579:1783–8. [PubMed: 15763552]
24. Dobrzynski M, Rodríguez JV, Kaandorp J. a, Blom JG. Computational methods for diffusion-influenced biochemical reactions. Bioinformatics. 2007; 23:1969–77. [PubMed: 17537752]
25. Costanzo M, et al. The genetic landscape of a cell. Science. 2010; 327:425–31. [PubMed: 20093466]
26. Rual J-F, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005; 437:1173–8. [PubMed: 16189514]
27. Stelzl U, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005; 122:957–68. [PubMed: 16169070]
28. Muñoz Descalzo S, et al. A competitive protein interaction network buffers Oct4-mediated differentiation to promote pluripotency in embryonic stem cells. Mol. Syst. Biol. 2013; 9:694. [PubMed: 24104477]
29. Xu H, Ang Y-S, Sevilla A, Lemischka IR, Ma'ayan A. Construction and Validation of a Regulatory Network for Pluripotency and Self-Renewal of Mouse Embryonic Stem Cells. PLoS Comput. Biol. 2014; 10:e1003777. [PubMed: 25122140]
30. Yu J, et al. Human induced pluripotent stem cells free of vector and transgene sequences. Science. 2009; 324:797–801. [PubMed: 19325077]
31. Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008; 132:1049–61. [PubMed: 18358816]
32. Boyer, L. a, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–56. [PubMed: 16153702]
33. Gohlke JM, et al. Characterization of the proneural gene regulatory network during mouse telencephalon development. BMC Biol. 2008; 6:15. [PubMed: 18377642]
34. Deneris ES, Wyler SC. Serotonergic transcriptional networks and potential importance to mental health. Nat. Neurosci. 2012; 15:519–27. [PubMed: 22366757]
35. Dagley, S.; Nicholson, D. Introduction to metabolic pathways. Blackwell Science Ltd; 1970.
36. Michal, G. Biochemical Pathways. Wiley-Blackwell; 1999.

37. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012; 40:D109–14. [PubMed: 22080510]
38. Croft D, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014; 42:D472–7. [PubMed: 24243840]
39. Kohn K. Functional capabilities of molecular network components controlling the mammalian G1/S cell cycle phase transition. Oncogene. 1998; 16:1065–1075. [PubMed: 9519880]
40. Le Novère N, et al. The Systems Biology Graphical Notation. Nat. Biotechnol. 2009; 27:735–741. [PubMed: 19668183] SBGN is a set of standard graphical languages to describe biological pathways. Akin to the electrical circuit standards, its use allows to interpret maps without the need of external information and legend.
41. Tozluoğlu M, Karaca E, Haliloglu T, Nussinov R. Cataloging and organizing p73 interactions in cell cycle arrest and apoptosis. Nucleic Acids Res. 2008; 36:5033–49. [PubMed: 18660513]
42. Kohn K. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Mol Biol Cell. 1999; 10:2703–2734. [PubMed: 10436023]
43. Pommier Y, Sordet O, Antony S, Hayward RL, Kohn KW. Apoptosis defects and chemotherapy resistance: molecular interaction maps and networks. Oncogene. 2004; 23:2934–49. [PubMed: 15077155]
44. Hlavacek WS, et al. Rules for modeling signal-transduction systems. Sci. STKE. 2006; 2006:re6. [PubMed: 16849649]
45. Danos, V.; Feret, J.; Fontana, W.; Harmer, R.; Krivine, J.; Caires, L.; Vasconcelos, V. Rule-based modelling of cellular signalling. in CONCUR 2007 - LNCS 4703. Springer-Verlag; 2007. p. 17-41.
46. Lopez CF, Muhlich JL, Bachman J. a, Sorger PK. Programming biological models in Python using PySB. Mol. Syst. Biol. 2013; 9:646. [PubMed: 23423320]
47. Bidkhori G, Moeini A, Masoudi-Nejad A. Modeling of tumor progression in NSCLC and intrinsic resistance to TKI in loss of PTEN expression. PLoS One. 2012; 7:e48004. [PubMed: 23133538]
48. Kiyatkin A, et al. Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops. J. Biol. Chem. 2006; 281:19925–38. [PubMed: 16687399]
49. Ung CY, et al. Simulation of the regulation of EGFR endocytosis and EGFR-ERK signaling by endophilin-mediated RhoA-EGFR crosstalk. FEBS Lett. 2008; 582:2283–90. [PubMed: 18505685]
50. Yamada S, Shiono S, Joo A, Yoshimura A. Control mechanism of JAK/STAT signal transduction pathway. FEBS Lett. 2003; 534:190–196. [PubMed: 12527385]
51. Hucka M, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003; 19:524–531. [PubMed: 12611808] SBML was a game-changing tool, allowing modellers to exchange and reuse models in different software without rewriting them from scratch. Moreover, its explicit semantic allowed the development of new approaches to process, analyse and enrich models.
52. Fernández-Suárez XM, Rigden DJ, Galperin MY. The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. Nucleic Acids Res. 2014; 42:D1–6. [PubMed: 24316579]
53. Orchard S, et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014; 42:D358–63. [PubMed: 24234451]
54. Franceschini A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41:D808–15. [PubMed: 23203871]
55. Caspi R, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2014; 42:D459–71. [PubMed: 24225315]
56. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013; 41:D793–800. [PubMed: 23143270]
57. Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. Nucleic Acids Res. 2006; 34:D504–6. [PubMed: 16381921]

58. Wittig U, et al. SABIO-RK--database for biochemical reaction kinetics. Nucleic Acids Res. 2012; 40:D790–6. [PubMed: 22102587]
59. Scheer M, et al. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 2011; 39:D670–6. [PubMed: 21062828]
60. Büchel F, et al. Path2Models: large-scale generation of computational models from biochemical pathway maps. BMC Syst. Biol. 2013; 7:116. [PubMed: 24180668]
61. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D. How to infer gene networks from expression profiles. Mol. Syst. Biol. 2007; 3:78. [PubMed: 17299415]
62. Hurley D, et al. Gene network inference and visualization tools for biologists: application to new human transcriptome datasets. Nucleic Acids Res. 2012; 40:2377–98. [PubMed: 22121215]
63. Chang G, et al. High-throughput sequencing reveals the disruption of methylation of imprinted gene in induced pluripotent stem cells. Cell Res. 2014; 24:293–306. [PubMed: 24381111]
64. Villaverde AF, Banga JR. Reverse engineering and identification in systems biology : strategies , perspectives and challenges. 2014; 11:20130505.Review of the various aspects of reverse engineering used to build models, including network inference, model identifiability and parameter estimation, taken from different points of views.
65. De Smet R, Marchal K. Advantages and limitations of current network inference methods. Nat. Rev. Microbiol. 2010; 8:717–29. [PubMed: 20805835]
66. He F, Balling R, Zeng A-P. Reverse engineering and verification of gene networks: principles, assumptions, and limitations of present methods and future perspectives. J. Biotechnol. 2009; 144:190–203. [PubMed: 19631244]
67. Haury A, Mordelet F, Vera-licona P, Vert J. TIGRESS : Trustful Inference of Gene REgulation using Stability Selection. BMC Syst. Biol. 2012; 6:145. [PubMed: 23173819]
68. Dunn S-J, Martello G, Yordanov B, Emmott S, Smith a. G. Defining an essential transcription factor program for naive pluripotency. Science (80-.). 2014; 344:1156–1160. [PubMed: 24904165]
69. Faith JJ, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 2007; 5:e8. [PubMed: 17214507]
70. Margolin, A. a, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006; 7(Suppl 1):S7. [PubMed: 16723010]
71. Gitter A, Klein-Seetharaman J, Gupta A, Bar-Joseph Z. Discovering pathways by orienting edges in protein interaction networks. Nucleic Acids Res. 2011; 39:e22. [PubMed: 21109539]
72. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J. Comput. Biol. 2000; 7:601–20. [PubMed: 11108481]
73. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics. 2004; 20:3594–603. [PubMed: 15284094]
74. Balov, N.; Salzman, P. CatNet: Categorical Bayesian Network Inference. 2011. at <http://cran.r-project.org/web/packages/catnet/>
75. Sachs K, Perez O, Pe'er D, Lauffenburger D. a, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science. 2005; 308:523–9. [PubMed: 15845847]
76. Feizi S, Marbach D, Médard M, Kellis M. Network deconvolution as a general method to distinguish direct dependencies in networks. Nat. Biotechnol. 2013; 31:726–33. [PubMed: 23851448]
77. Nelander S, et al. Models from experiments: combinatorial drug perturbations of cancer cells. Mol. Syst. Biol. 2008; 4:216. [PubMed: 18766176]
78. Bonneau R, et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol. 2006; 7:R36. [PubMed: 16686963]
79. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. Science. 2003; 301:102–5. [PubMed: 12843395]
80. Jaeger J, et al. Dynamic control of positional information in the early Drosophila embryo. 2004; 430:2–5.

81. Cantone I, et al. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. Cell. 2009; 137:172–81. [PubMed: 19327819]
82. Marbach D, et al. Wisdom of crowds for robust gene network inference. Nat. Methods. 2012; 9:796–804. [PubMed: 22796662] Results of the network inference challenge of the DREAM5 competition. While within each type of approach some tools perform better than others, the overall best result is obtained by using a combination of several approaches.
83. Hurley DG, et al. NAIL, a software toolset for inferring, analyzing and visualizing regulatory networks. Bioinformatics. in the pre,
84. Allen, JP. Biophys. Chem. Wiley-Blackwell; 2008. p. 134-162.
85. Le Novère, N.; Endler, L.; Schneider, MV. Silico Syst. Biol. Methods Mol. Biol. Vol. 1021. Humana Press; 2013. Using Chemical Kinetics to Model Biochemical Pathways; p. 147-167.1021
86. Keller R, et al. The systems biology simulation core algorithm. BMC Syst. Biol. 2013; 7:55. [PubMed: 23826941]
87. Hoops S, et al. COPASI--a COmplex PAthway SImulator. Bioinformatics. 2006; 22:3067–74. [PubMed: 17032683]
88. Polynikis A, Hogan SJ, Bernardo M. Comparing different ODE modelling approaches for gene regulatory networks. J. Theor. Biol. 2009; 261:511–530. [PubMed: 19665034] Systematic and quantitative comparison of different ODEs methods used to model gene regulatory networks. The work presents the underlying hypothesis, as well as advantages and shortcomings.
89. Dalle Pezze P, et al. A dynamic network model of mTOR signaling reveals TSC-independent mTORC2 regulation. Sci. Signal. 2012; 5:ra25. [PubMed: 22457331]
90. Nelson D, et al. Oscillations in NF-kappaB signaling control the dynamics of gene expression. Science (80-.). 2004; 306:704–708. [PubMed: 15499023]
91. Ashall L, et al. Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. Science (80-.). 2009; 324:242–246. [PubMed: 19359585]
92. Bergethon, P. r. Phys. basis Biochem. Springer; 1998. p. 480-497.
93. Huang CY, Ferrell JE. Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc. Natl. Acad. Sci. U. S. A. 1996; 93:10078–83. [PubMed: 8816754]
94. Chen, WW.; Niepel, M.; Sorger, PK. Classic and contemporary approaches to modeling biochemical reactions; 2010. p. 1861-1875.doi:10.1101/gad.1945410.Freely
95. Segel, IH. Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems. Wiley-Blackwell; 1993.
96. Cornish-Bowden, A. Fundamentals of Enzyme Kinetics. Wiley VCH; 2012.
97. Voit EO. Biochemical Systems Theory: A Review. ISRN Biomath. 2013; 2013:1–53.
98. Parthimos D, Haddock RE, Hill CE, Griffith TM. Dynamics of a three-variable nonlinear model of vasomotion: comparison of theory and experiment. Biophys. J. 2007; 93:1534–56. [PubMed: 17483163]
99. Yao G, Tan C, West M, Nevins JR, You L. Origin of bistability underlying mammalian cell cycle entry. Mol. Syst. Biol. 2011; 7:485. [PubMed: 21525871]
100. Novák B, Tyson JJ. Design principles of biochemical oscillators. Nat. Rev. Mol. Cell Biol. 2008; 9:981–91. [PubMed: 18971947]
101. Goldbeter A, Pourquié O. Modeling the segmentation clock as a network of coupled oscillations in the Notch, Wnt and FGF signaling pathways. J. Theor. Biol. 2008; 252:574–85. [PubMed: 18308339]
102. Ozbudak EM, Lewis J. Notch signalling synchronizes the zebrafish segmentation clock but is not needed to create somite boundaries. PLoS Genet. 2008; 4:e15. [PubMed: 18248098]
103. Glass L, Kauffman S. The logical analysis of continuous, non-linear biochemical control networks. J. Theor. Biol. 1973; 39:129–103.
104. Coutinho R, Fernandez B, Lima R, Meyroneinc A. Discrete time piecewise affine models of genetic regulatory networks. J. Math. …. 2006; 52:524–570.
105. De Jong H, et al. Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bull. Math. Biol. 2004; 66:301–40. [PubMed: 14871568]

106. De Jong H, Geiselmann J, Hernandez C, Page M. Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. Bioinformatics. 2003; 19:336–344. [PubMed: 12584118]
107. Moles CG, Mendes P, Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res. 2003; 13:2467–74. [PubMed: 14559783]
108. Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger D. a. Logic-based models for the analysis of cell signaling networks. Biochemistry. 2010; 49:3216–24. [PubMed: 20225868] A good introduction to logic modelling, including its different variants and available software tools.
109. Sánchez L, Van Helden J, Thieffry D. Establishment of the Dorso-ventral Pattern During Embryonic Development of Drosophila melanogaster : a Logical Analysis. J. Theor. Biol. 1997; 189:377–389. [PubMed: 9446747]
110. Yuh C-H, Bolouri H, Davidson EH. Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. Science (80-.). 1998; 279:1896–1902. [PubMed: 9506933]
111. Bonzanni N, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. Bioinformatics. 2013; 29:i80–8. [PubMed: 23813012]
112. Wittmann DM, et al. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. BMC Syst. Biol. 2009; 3:98. [PubMed: 19785753]
113. Calzone L, et al. Mathematical modelling of cell-fate decision in response to death receptor engagement. PLoS Comput. Biol. 2010; 6:e1000702. [PubMed: 20221256]
114. Grieco L, et al. Integrative modelling of the influence of MAPK network on cancer cell fate decision. PLoS Comput. Biol. 2013; 9:e1003286. [PubMed: 24250280]
115. Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G. Synchronous versus asynchronous modeling of gene regulatory networks. Bioinformatics. 2008; 24:1917–25. [PubMed: 18614585]
116. Ahmad J, Bernot G, Comet J-P, Lime D, Roux O. Hybrid Modelling and Dynamical Analysis of Gene Regulatory Networks with Delays. Complexus. 2006; 3:231–251.
117. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics. 2002; 18:261–74. [PubMed: 11847074]
118. Vahedi G, Faryabi B, Chamberland J-F, Datta A, Dougherty ER. Sampling-rate-dependent probabilistic Boolean networks. J. Theor. Biol. 2009; 261:540–7. [PubMed: 19716832]
119. Liang J, Han J. Stochastic Boolean networks: an efficient approach to modeling gene regulatory networks. BMC Syst. Biol. 2012; 6:113. [PubMed: 22929591]
120. Helikar T, Kochi N, Konvalina J, Rogers JA. Boolean Modeling of Biochemical Networks. Open Bioinforma. J. 2011; 5:16–25.
121. Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger D. a. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. PLoS Comput. Biol. 2009; 5:e1000340. [PubMed: 19343194]
122. Terfve C, et al. CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. BMC Syst. Biol. 2012; 6:133. [PubMed: 23079107]
123. MacNamara A, Terfve C. State–time spectrum of signal transduction logic models. Phys. Biol. 2012; 9:045003. [PubMed: 22871648]
124. Carrillo M, Góngora P. a, Rosenblueth D. a. An overview of existing modeling tools making use of model checking in the analysis of biochemical networks. Front. Plant Sci. 2012; 3:155. [PubMed: 22833747]
125. Jusko WJ. Moving from basic toward systems pharmacodynamic models. J. Pharm. Sci. 2013; 102:2930–40. [PubMed: 23681608]
126. Takahashi K, Kaizu K, Hu B, Tomita M. A multi-algorithm, multi-timescale method for cell simulation. Bioinformatics. 2004; 20:546–538.
127. McAdams H, Shapiro L. Circuit simulation of genetic networks. Sci. (New York, NY). 1995; 269:656–650.
128. Singhania R, Sramkoski RM, Jacobberger JW, Tyson JJ. A hybrid model of mammalian cell cycle regulation. PLoS Comput. Biol. 2011; 7:e1001077. [PubMed: 21347318]

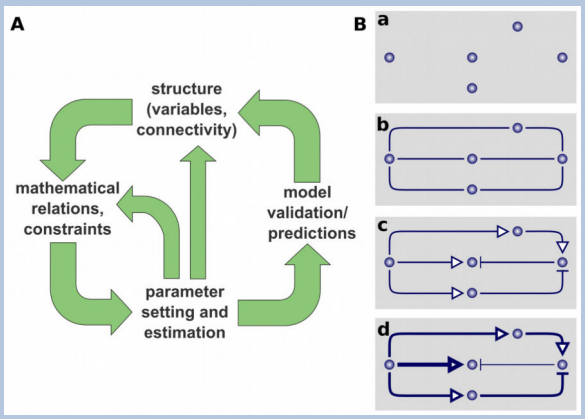
129. Ryll, a, et al. A model integration approach linking signalling and gene-regulatory logic with kinetic metabolic models. Biosystems. 2014; 124:26–38. [PubMed: 25063553]
130. Dmem H, Chen J, Covert MW, Knight EM, Reed JL. Integrating high-throughput and computational data. 2004; 429:2–6.
131. Herrgård, MJ.; Lee, B.; Portnoy, V.; Palsson, BØ. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae; 2006. p. 627-635.doi:10.1101/gr.4083206.predict
132. Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol. Syst. Biol. 2007; 3:101. [PubMed: 17437026]
133. Mattioni M, Le Novère N. Integration of biochemical and electrical signaling-multiscale model of the medium spiny neuron of the striatum. PLoS One. 2013; 8:e66811. [PubMed: 23843966]
134. Klionsky DJ, et al. Guidelines for the use and interpretation of assays for monitoring autophagy. Autophagy. 2012; 8:445–544. [PubMed: 22966490]
135. Bhalla U, Iyengar R. Emergent properties of networks of biological signaling pathways. Science (80-.). 1999; 283:381–387. [PubMed: 9888852]
136. Stefan MI, Edelstein SJ, Le Novère N. An allosteric model of calmodulin explains differential activation of PP2B and CaMKII. Proc Natl Acad Sci U S A. 2008; 105:10768–10773. [PubMed: 18669651]
137. Le Novère N, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 2005; 23:1509–15. [PubMed: 16333295]
138. Waltemath D, et al. Minimum Information About a Simulation Experiment (MIASE). PLoS Comput. Biol. 2011; 7:e1001122. [PubMed: 21552546]
139. Waltemath D, et al. Reproducible computational biology experiments with SED-ML--the Simulation Experiment Description Markup Language. BMC Syst. Biol. 2011; 5:198. [PubMed: 22172142]
140. François P, Hakim V. Core genetic module: The mixed feedback loop. Phys. Rev. E. 2005; 72:031908.
141. Naldi, a, et al. Logical modelling of regulatory networks with GINsim 2.3. Biosystems. 2009; 97:134–9. [PubMed: 19426782]
142. Calzone L, Fages F, Soliman S. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. Bioinformatics. 2006; 22:1805–7. [PubMed: 16672256]
143. Funahashi A, Morohashi M, Kitano H, Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. BIOSILICO. 2003; 1:162–159.
144. Myers CJ, et al. iBioSim: a tool for the analysis and design of genetic circuits. Bioinformatics. 2009; 25:2848–9. [PubMed: 19628507]
145. Ermentrout, B.; Le Novère, N. Comput. Syst. Neurobiol. Springer; 2012. p. 519-531.
146. Müssel C, Hopfensitz M, Kestler H. a. BoolNet - an R package for generation, reconstruction, and analysis of Boolean networks. Bioinformatics. 2010; 26:1378–1380. [PubMed: 20378558]

Europe PMC Funders Author Manuscripts

Box 1

Building mathematical models of biological processes

Far from being a linear activity, the process of building a mathematical model is a cycle of multiple iterations in which the appropriate number of variables, the mathematical relationships and the parameter values are selected, and in which numerical simulations and other mathematical analyses are performed to both reproduce observations and form predictions (see the figure, part A). In each cycle, the model is extended to include novel variables that are necessary to account for observed behaviours and is also pruned of superfluous complexity. The most sophisticated simulation approaches are of little use if the initial model structure contains basic flaws.



Model building is a layered procedure, and new biological insights can be obtained at each stage.

The first layer is to determine, or infer from experimental data sets, the biological entities (represented as blue circles) to be represented in the model: that is, which genes or which molecular species will be part of the network (see the figure, part Ba). The number of entities to include depends on both the question asked and the data available to parameterize and validate the model. A model does not necessarily need to include all that is known about a system. Biological processes and structures can be described at different levels of granularity (for example, different models have been constructed using 1 (133) , 4 (134) or 32 (135) states to represent calcium–calmodulin). Choosing when to be biochemically accurate and when to be approximate is one of the most challenging steps in the model-building process. Selecting too many molecular species might lead to difficulties when building the mathematical model. The next step constitutes searching for possible interactions between the components (see the figure, part Bb), which can be added, for instance, from functional genomics experiments. The analysis of such a network can already provide information at the level of the system (21). A deeper description includes the directionality of the relationships, transforming the network into a pathway, and permits the description of the flux of information in the network (see the figure, part Bc). Finally, the relationships can be characterized and quantified (see the figure, part Bd; (line thickness represents the strength of influences).

A mathematical model of a system is made of three structures. First, the variables correspond to biological entities, the activity or quantity of which we know or want to determine. Variables can represent physical constituents (such as pools of molecules) or parameters (such as kinetic and equilibrium constants or characteristic dimensions) that are either constant or varying during the simulation. Second, the mathematical relationships link the variables together and represent what we already know or what we want to test. Mathematical relationships come in many guises: for instance, assignments, rates of change, sampling and logic rules. Third, the constraints represent the context of the analysis or represent processes that we choose to ignore in the project. Examples of simple constraints are concentrations that cannot be negative and conservation laws. An important set of constraints is the initial conditions: for example, the values taken by all variables at the beginning of a simulation.

All steps of the model building and simulation procedures must be carefully documented to enable verification and reproduction. The computational systems biology community has developed sets of guidelines that list all information that must be shared together with the model — namely, Minimal Information Required for the Annotation of Models (MIRIAM) (136) and Minimal Information about a Simulation Experiment (MIASE) (137). The required information can be encoded in standard formats: for instance, Systems Biology Markup Language (SBML) (51) for the models and Simulation Experiment Description Markup Language (SED-ML) (138) for the simulations and analyses. These open standards have had an important impact on the field of modelling in systems biology, opening the door to model sharing and reuse, as well as automated model building and analyses (60).

Europe PMC Funders Author Manuscripts

Box 2

Quantitative modelling based on chemical kinetics

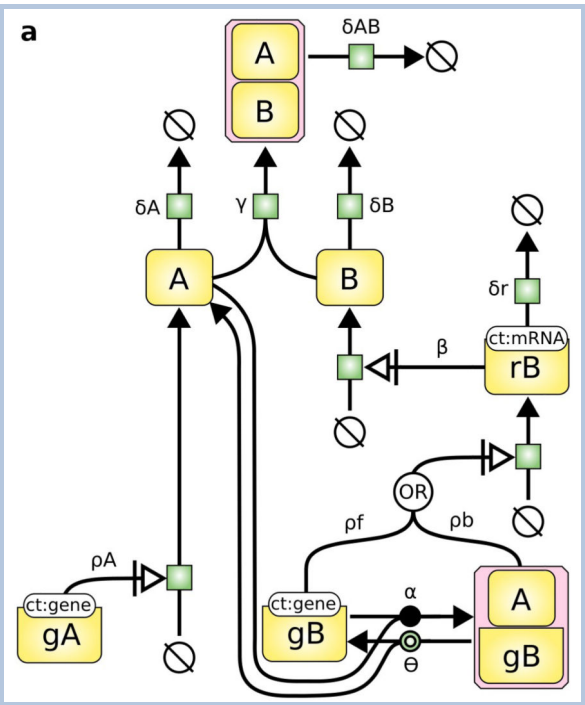
Chemical kinetics is based on the notion of processes that consume reactants and generate products. The rate of a process is determined by the quantity of the components involved, their REACTION ORDER and parameters. The exact form of the mathematical expression depends on the assumptions made. The behaviour of a system emerges from the combination of all of the processes affecting its components. The value of each variable — for instance, the amount of a protein or the activity of a gene — is positively or negatively affected by processes such as binding or catalysis. The resulting effect on a given component is the sum of the rates of all processes that the component is involved in, multiplied by its stoichiometries for these processes. For instance, if two molecules of X reversibly bind to form a molecule Y, the stoichiometries for X are −2 for the formation of Y and +2 for its dissociation, whereas the stoichiometries for Y are 1 for its formation and −1 for its dissociation. If we model the processes using mass action law, the reaction orders are 2 for the formation of Y (2 molecules of X are needed to form 1 Y) and 1 for its dissociation (1 molecule of Y dissociates into 2 X). Therefore, equations describing the temporal evolution of X and Y concentrations are as shown below, where $rate_{ass}$ and $rate_{diss}$ are the rates of association and dissociation, respectively, and k_{ass} and k_{diss} are the association and dissociation constants, respectively.

$$\mathrm{d}[\mathrm{X}]/\mathrm{dt} = -2 \cdot rate_{ass} + 2 \cdot rate_{diss} = -2 \cdot k_{ass} \cdot [\mathrm{X}]2 + 2 \cdot k_{diss} \cdot [\mathrm{Y}]$$
$$\mathrm{d}[\mathrm{Y}]/\mathrm{dt} = +1 \cdot rate_{ass} - 1 \cdot rate_{diss} = +1 \cdot k_{ass} \cdot [\mathrm{X}]2 - 1 \cdot k_{diss} \cdot [\mathrm{Y}]$$

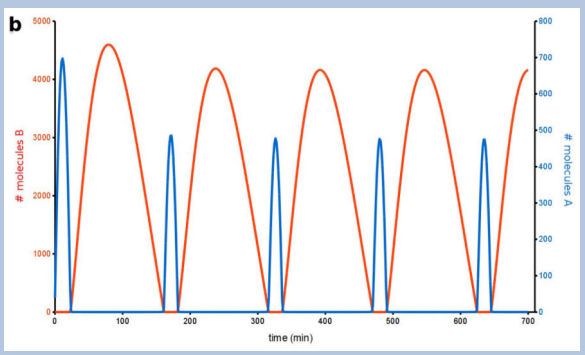
The addition of simple processes allows the rapid construction of more-complex systems. Part **a** of the figure represents, using the Systems Biology Graphical Notation (SBGN) process description language (40), a simple two-gene system that can display different behaviours, such as monostability and bistability or oscillations, depending on the parameter values (139). We can model the system's behaviour by following the state of gene B (reversibly binding to protein A), the amount of mRNA B (increased by gene B expression and decreased by degradation), the amount of protein A (increased by gene A expression and decreased by binding to gene B, binding to protein B and degradation), protein B (increased by mRNA B expression and decreased by binding to protein A and degradation) and complex AB (increased by the binding of A to B and decreased by degradation). The resulting equations are as follows (see the figure, part **a**).

$$\mathrm{d}[\mathrm{gB}]/\mathrm{dt} = \theta \cdot (1 - \mathrm{gB}) - \alpha \cdot \mathrm{gB} \cdot [\mathrm{A}]$$
$$\mathrm{d}[\mathrm{rB}]/\mathrm{dt} = \rho_{\mathrm{f}} \cdot \mathrm{gB} + \rho_{\mathrm{b}} \cdot (1 - \mathrm{gB}) - \delta_{\mathrm{r}} \cdot [\mathrm{rB}]$$
$$\mathrm{d}[\mathrm{A}]/\mathrm{dt} = (1 \cdot)\,\rho_{\mathrm{A}} + \theta \cdot (1 - \mathrm{gB}) - \alpha \cdot \mathrm{gB} \cdot [\mathrm{A}] - \gamma \cdot [\mathrm{A}] \cdot [\mathrm{B}] - \delta_{\mathrm{A}} \cdot [\mathrm{A}]$$
$$\mathrm{d}[\mathrm{B}]/\mathrm{dt} = \beta \cdot [\mathrm{rB}] - \gamma \cdot [\mathrm{A}] \cdot [\mathrm{B}] - \delta_{\mathrm{B}} \cdot [\mathrm{B}]$$
$$\mathrm{d}[\mathrm{AB}]/\mathrm{dt} = \gamma \cdot [\mathrm{A}] \cdot [\mathrm{B}] - \delta_{AB} \cdot [\mathrm{AB}]$$

Europe PMC Funders Author Manuscripts



Using a biochemical simulator such as COPASI (86), one does not need to write the differential equations. The modeller writes down the chemical equations and the rates of the reactions. The software will then numerically solve the resulting equations (see the figure, part **b**). Alternatively, one can run a stochastic simulation, in which case each process is considered separately. The model and simulation description can be found in Systems Biology Markup Language (SBML) and Simulation Experiment Description Markup Language (SED-ML), and in COPASI format in the BioModels database (accession number: BIOMD0000000539).



Europe PMC Funders Author Manuscripts

Box 3

Logic modelling

Logic modelling is based on the idea that a variable can take a discrete number of states or values (two in the case of Boolean models) and that the state of a variable is decided by a logical combination of the states of other variables. The system can be updated synchronously, with the values of all variables being calculated after a transition, or asynchronously, when variables undergo transitions one at a time (114).

We can create a logic version of the system presented in BOX 2 by building a model with three nodes representing protein A, protein B and the complex AB (see the figure, part **a**). The activity of A is represented by a Boolean variable. It is inhibited by the complex AB, otherwise it is always ‘on’. The activity of AB is represented by a Boolean variable and is stimulated if both A and B are active. Finally, the activity of B is represented by three values. It can be off, low or high if A is on (and stimulates its production) and AB is off. Note that in the following expressions, B is true if B = 1 or B = 2.

A = 0 if AB

A = 1 if not AB

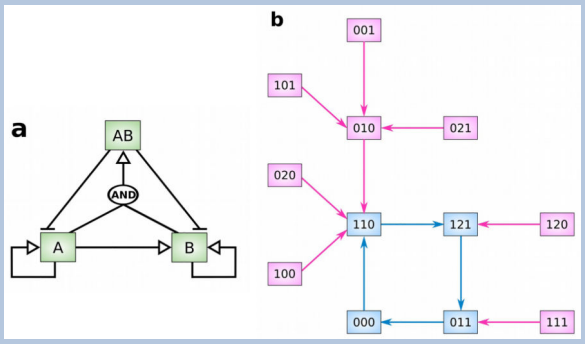
B = 0 if not A and AB

B = 1 if (not A and not AB) or (A and AB)

B = 2 if A and not AB

AB = 0 if not A or not B

AB = 1 if A and B



The model was implemented using the GINsim software (140). The synchronous simulation of the logic rules permits the tracing of trajectories across the ensemble of states. The combination of all of these trajectories forms the state-transition graph (see the figure, part **b**). Whatever the starting state, the system will end up as a circular attractor in which all three variables oscillate.

Europe PMC Funders Author Manuscripts

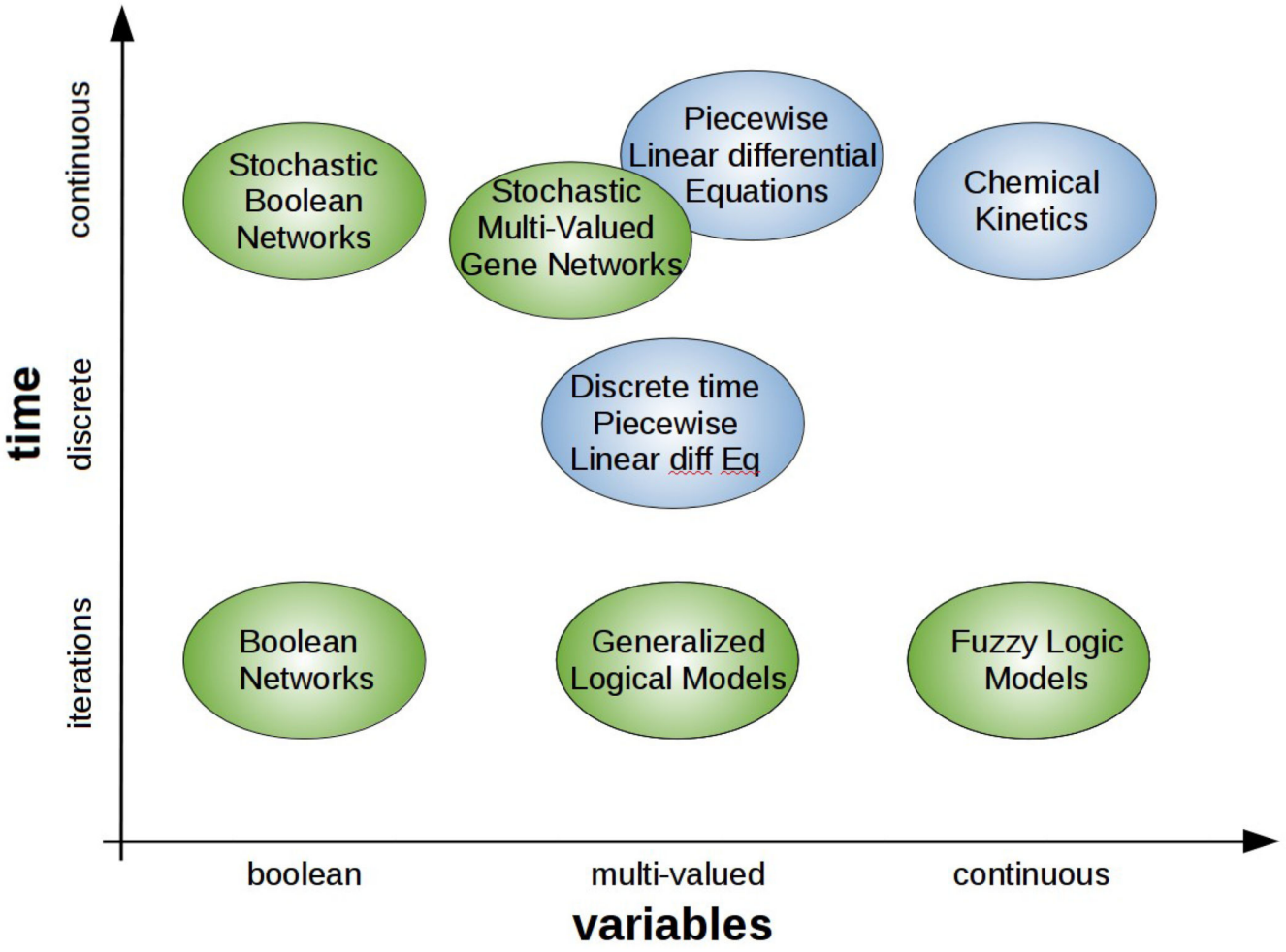


Figure 1. Granularity of time representation and variable values for various modelling approaches

Variables in a model can take unbounded values (for example, concentrations or the number of molecules), multiple although limited values (for example, null, low, medium or high) or Boolean values (present or absent, or active or inactive). Progression of the variables during simulations can be represented using continuous time (mirroring the real world) in a discrete manner (with updates made after specified time durations), or using iterations (which do not necessarily represent any specific duration). Green methods are updated according to logic rules, whereas purple methods compute the new values of variables using quantitative mathematics.

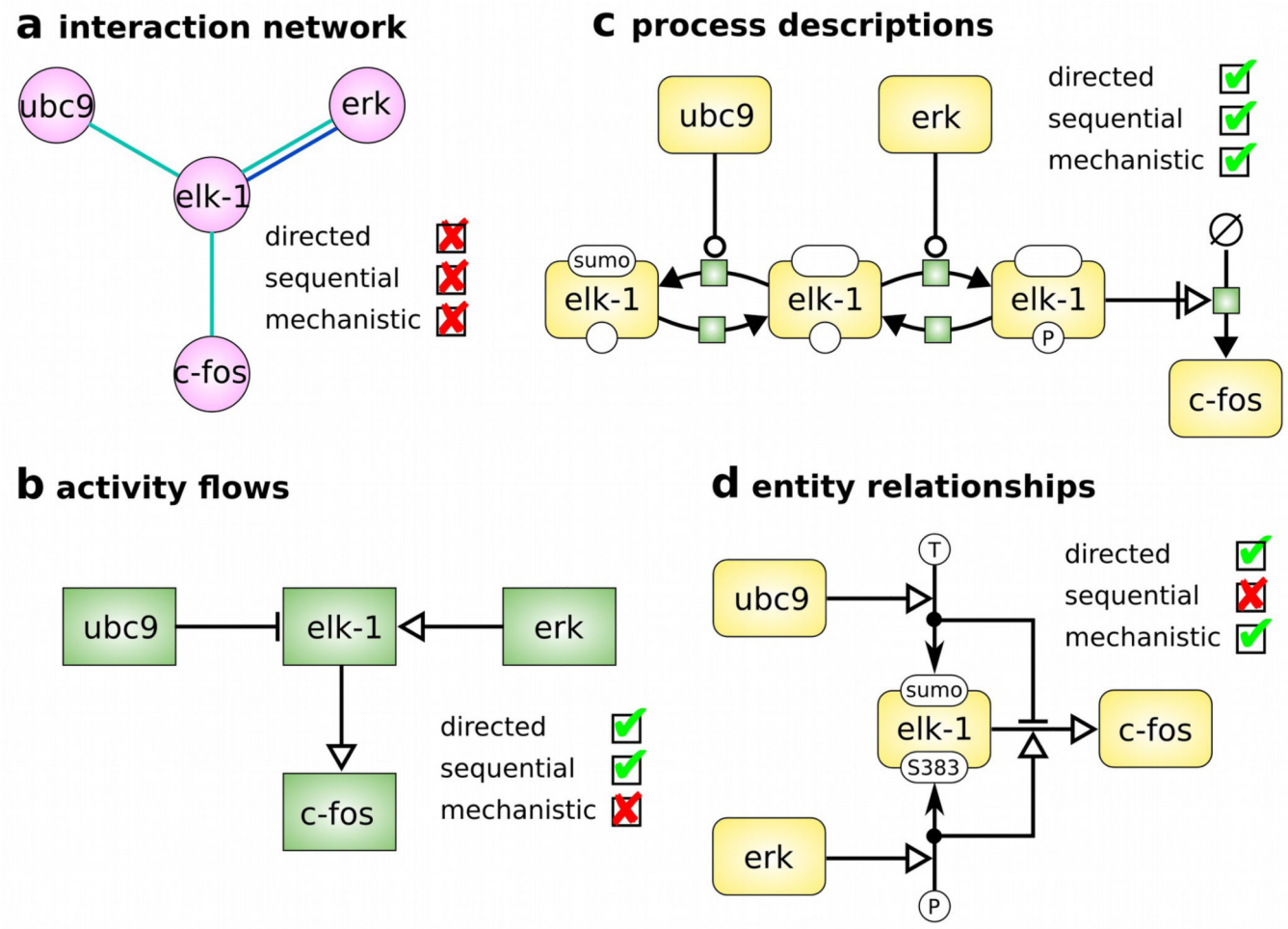


Figure 2. The four views of systems biology

Four different types of networks used to represent biological processes and their features are shown. a | An interaction network can be used to represent physical interactions (black line) — such as that between extracellular signal-regulated kinase (ERK) and ELK1 — and functional interactions (grey lines), such as those between UBC9 (also known as UBE2I), ERK, ELK1 and c-FOS. b | An activity flow can be used to show the stimulation of c-FOS activity by ELK1 activity, the stimulation of ELK1 activity by ERK activity, and its inhibition by UBC9 activity. c | A detailed process description can be used to show the catalysis of ELK1 sumoylation (SUMO) and phosphorylation (P), their reversed reactions, and the trigger of c-FOS expression. The graph is simplified by the inexistence of ELK1 with both covalent modifications. d | Entity relationships can be used to describe the stimulation of sumoylation and phosphorylation of ELK1 by UBC9 and ERK, respectively, and the influence of these processes on c-FOS.

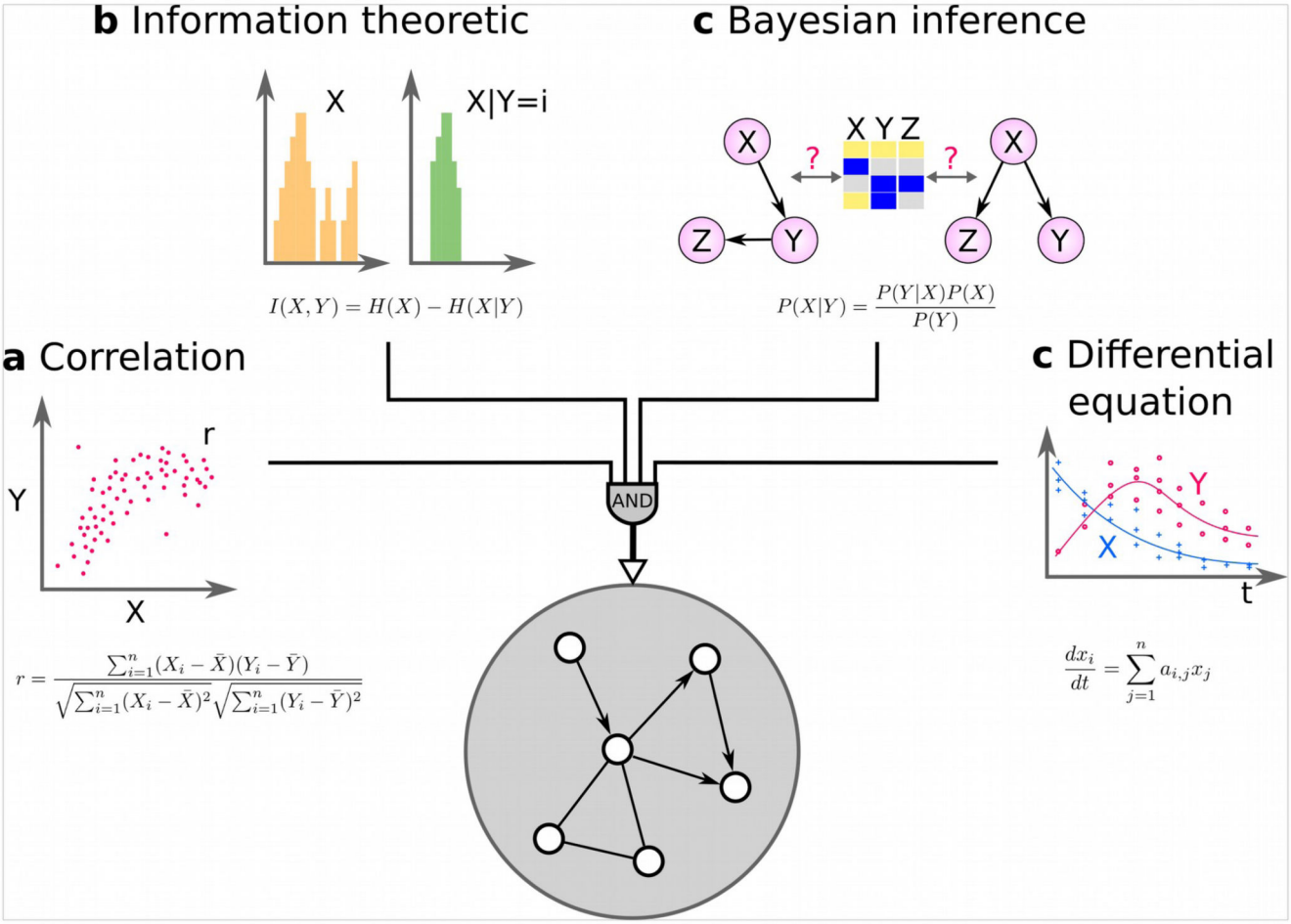


Figure 3. Network inference methods

The four main approaches to infer networks from data include: correlation (part a), information theoretic (part b), Bayesian inference (part c) and differential equations (part d). The combination of several approaches seems to be the most robust method to obtain the correct network.

Europe PMC Funders Author Manuscripts

Table 1

Comparison matrix of quantitative and qualitative models

	Quantitative model	**Logic model**
Suitable for	Time series	Phenotypes
Time representation	Linear representation	Abstract iterations
Variables	Quantitative	Qualitative
Mechanism representation	Yes	No
What can we do?	Compute concentrations and durations; evaluate the effect of parameter values	Compute state transitions and attractors (steady-states and cyclic attractors)
Data necessary to build the model	Molecular species, genes, interactions, biochemical processes	Activities, defined phenotypes, rules linking those
Data to parameterize and validate the model	Amount of molecular species, timecourses, quantitative phenotype	Perturbations of activities such as RNA interference, inhibitors, qualitative phenotypes
Advantages	Quantitative, precise; direct comparison with quantitative measurements; large existing toolkit	Easy to build; easy to compose; easy simulation of perturbations
Weaknesses	Requires quantitative knowledge of initial conditions and kinetics	Cannot provide quantitative predictions; difficult to choose between alternative behaviours

Europe PMC Funders Author Manuscripts

Table 2

Example of freely available software used to build and analyse models

Name	Features	License	Refs
Gene regulatory network inference			
ARACNE	Information theoretic	Non-commercial license	70
BANJO	Bayesian inferences	Non-commercial license	73
CatNet	Bayesian inferences	General Public License	-
Inferelator	ODEs	No license	77
NAIL	Multiple	Apache License	82
NIR	ODEs	Non-commercial license	78
TIGRESS	Regression	General Public License	67
Quantitative kinetic modelling			
BIOCHAM	ODEs	General Public License	141
CellDesigner	ODEs, stochastic	Gratis	142
COPASI	ODEs, stochastic	Artistic License	86
DBSolve	ODE	Gratis	-
E-Cell Project	ODEs, stochastic	General Public License	125
iBioSim	ODEs, stochastic	MIT License	143
SBML simulator	ODEs	Lesser General Public License	62
XPP-Aut	ODEs	General Public License	144
Qualitative modelling			
BoolNET	Logic models	Artistic License	145
CellNetOptimizer	Logic models	General Public License	121
GINsim	Logic models	General Public License	140
GNA	Piecewise linear equations	Gratis for non-profit academic research	105